NeoBiota 72: 183–203 (2022)
doi: 10.3897/neobiota.72.77266
https://neobiota.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

Modelling the likelihood of entry of marine non-indigenous species from internationally arriving vessels to maritime ports: a case study using New Zealand data

Rezvan Hatami[1], Graeme Inglis[2], Stephen E. Lane[1], Abraham Growcott[3], Daniel Kluza[3], Catherine Lubarsky[3], Charlotte Jones-Todd[2], Kimberley Seaward[2], Andrew P. Robinson[1]

1 *Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne, Parkville, Melbourne, Australia* **2** *National Institute of Water & Atmospheric Research Ltd (NIWA), Wellington, New Zealand* **3** *Biosecurity New Zealand, Ministry for Primary Industries, Wellington, New Zealand*

Corresponding author: Rezvan Hatami (Rezvan.hatami@hotmail.com)

Academic editor: Cascade Sorte | Received 29 October 2021 | Accepted 25 March 2022 | Published 18 April 2022

Citation: Hatami R, Inglis G, Lane SE, Growcott A, Kluza D, Lubarsky C, Jones-Todd C, Seaward K, Robinson AP (2022) Modelling the likelihood of entry of marine non-indigenous species from internationally arriving vessels to maritime ports: a case study using New Zealand data. NeoBiota 72: 183–203. https://doi.org/10.3897/neobiota.72.77266

Abstract

The establishment of marine non-indigenous species (NIS) in new locations can degrade environmental, socio-cultural, and economic values. Vessels arriving from international waters are the main pathway for the entry of marine NIS, via exposure due to ballast water discharge (hereafter, ballast discharge) and biofouling. We developed a systematic statistical likelihood-based methodology to investigate port-level marine NIS propagule pressure from ballast discharge and biofouling exposure using a combination of techniques, namely *k*-Nearest-Neighbour and random forest algorithms. Vessel characteristics and travel patterns were assessed as candidate predictors. For the ballast discharge analysis, the predictors used for model building were vessel type, dead weight tonnage, and the port of first arrival; the predictors used for the biofouling analysis were days since last antifouling paint, mean vessel speed, dead weight tonnage, and hull niche area. Propagule pressure for both pathways was calculated at a voyage, port and annual level, which were used to establish the relative entry score for each port. The model was applied to a case study for New Zealand. Biosecurity New Zealand has commissioned targeted marine surveillance at selected ports since 2002 to enable early detection of newly arrived marine NIS (Marine High-Risk Site Surveillance, MHRSS). The reported methodology was used to compare contemporary entry likelihoods between New Zealand ports. The results suggested that Tauranga now receives the highest volume of discharged ballast water and has the second most biofouling exposure compared to all other New Zealand ports. Auckland was predicted to receive the highest biofouling mass and was ranked tenth for ballast

Copyright *Rezvan Hatami et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

discharge exposure. Lyttelton, Napier, and New Plymouth also had a high relative ranking for these two pathways. The outputs from this study will inform the refinement of the MHRSS programme, facilitating continued early detection and cost-effective management to support New Zealand's wider marine biosecurity system. More generally, this paper develops an approach for using statistical models to estimate relative likelihoods of entry of marine NIS.

Keywords
ballast water, biofouling, biosecurity, marine ecosystems, non-indigenous species, vessels

Introduction

Marine ecosystems are vulnerable to the introduction of non-indigenous species (NIS) through human activities such as international trade and transportation, which have increased greatly in the last fifty years (Hulme 2009). International marine shipping, which contributes to 90% of world trade, is a leading source of introduction of many marine NIS (Hewitt et al. 2009; Hulme 2009; International Maritime Organization [IMO] 2017). Marine NIS have the potential to impact upon environmental, socio-cultural, and economic values. For example, in 1991, the outbreak of a virulent strain of cholera infected one million people and claimed thousands of lives and was linked to the discharge of ballast water in Peru (McCarthy and Khambaty 1994; Tauxe et al. 1995; Kolar and Lodge 2001). Therefore, it is essential to describe existing patterns of invasion and forecast future invasions for improving management of marine ecosystems (Drake and Lodge 2004).

The two main pathways associated with the spread of marine NIS via shipping are ballast water discharge (hereafter, ballast discharge) and biofouling exposure (Hewitt and Campbell 2010). Until recently, ballast water was thought to be the major contributor to marine NIS spread; however, the biofouling pathway has been assessed to be as great, if not a higher threat (Cordell et al. 2009; Hewitt and Campbell 2010; Bell et al. 2011). In New Zealand, it has been estimated that between 69–90% of established marine NIS are likely to have been introduced via biofouling, with ballast discharge being the second most important pathway (Cranfield et al. 1998; Kospartov et al. 2008). International and national policies have been implemented to minimise the spread of marine NIS through ballast water and biofouling, and New Zealand is currently the only country to have enacted mandatory biofouling regulations (Georgiades et al. 2020).

The use of predictive models can be a powerful tool for helping to understand the dynamics of invasion pathways through the exploration of factors that influence propagule pressure (Verling et al. 2005). Variables that influence propagule pressure can be classified as voyage properties (e.g., duration, source region, visited locations, arrival date or season), vessel characteristics (e.g., size, ballast water capacity, niche area type and extent) and behaviours (e.g., compliance with regulations), all of which influence

the likelihood of entry for both pathways (McGee et al. 2006; Lacoursière-Roussel et al. 2012; Minton et al. 2015).

Many studies have identified high-risk invasion pathways (Cope et al. 2015), global hotspots of invasions (Drake and Lodge 2004), possible source ports of invasion (Paini and Yemshanov 2012), and high-risk vessels arriving from international ports (Clarke et al. 2017). However, few studies have established how information about vessel traffic patterns and vessel characteristics for ballast water and biofouling pathways can be integrated to identify ports most likely to receive marine NIS. We demonstrate the use of an integrative modelling approach using the data available from New Zealand, to investigate propagule pressure differences among ports.

The Marine High-Risk Site Surveillance (**MHRSS**) Programme was established in 2002 to facilitate the early detection of selected marine NIS that were considered likely to impact on New Zealand's marine values. Currently, 11 ports and marinas are considered 'high-risk' for marine NIS entry and establishment (Seaward et al. 2015). The original site selection was based on proxy indicators using data from 1999, e.g., vessel visits and port habitat types (see Inglis 2001a, b, Inglis et al. 2006). Since the inception of this programme, changes in vessel movement patterns and behaviours are likely, and re-evaluation is appropriate. At the time of the study, the allocation of surveillance effort was identical to all ports except at Auckland, at which it was doubled.

The aim of this research is to develop a systematic statistical likelihood-based approach for predicting marine invasions. We investigate whether the ports and marinas currently surveyed as part of the MHRSS programme are most likely the locations where marine NIS will enter New Zealand with internationally arriving vessels via the ballast and biofouling pathways. We then assess whether the allocation of survey effort at each location should be modified based on these findings.

Methods

Datasets were provided by Biosecurity New Zealand (**BNZ**), the government agency responsible for New Zealand's biosecurity system. The main challenge in predicting the hazard presented by ballast discharge for the current case study is that contemporary ballast discharge records were not available. BNZ collected ballast discharge intention declaration data up until 2008, but risk analysts were concerned that the pattern and volume of transits may have changed substantially, hence the need to connect the historical and contemporary data described below.

The challenge in predicting the hazard due to biofouling exposure is to construct a model that permits prediction of the biofouling level based on easily captured vessel statistics. As described below, we used data from a historical biofouling study to construct this model.

For vessel ballast water and biofouling pathways we used two sets of historical and contemporary data as described below.

Historical data

Ballast discharge

A historical dataset of biofouling and ballast water declarations was compiled for vessels that visited New Zealand during 1998 to 2008, which contained records for 15,745 voyages by 2,607 vessels. This dataset contained information on the vessel's physical and voyage characteristics, and ballast discharge volumes detailed for each ballast tank. This dataset was examined to determine: (i) unique voyage and port visits by vessels and volumes of ballast discharged per vessel at each port, and (ii) management of the ballast and intention to discharge in New Zealand. The response variables used for ballast water modelling were the port of discharge and discharge volume per port. Vessel type, ballast capacity or Deadweight Tonnage (**DWT**), arrival port, and the intention of discharge were candidate predictor variables for estimating how much ballast water was discharged at each port. Arrival port is the first New Zealand port a vessel encounters in a voyage. Many voyages declared more than one port of ballast discharge, i.e., each journey has a first arrival port, but the vessel might visit other ports as part of the same voyage and discharge ballast. For example, a vessel's voyage to New Zealand may include visits to Whangarei, Tauranga and Auckland, with intended ballast discharge at Whangarei and Tauranga but not Auckland; here the recorded intended ballast discharge volumes for the first two ports (i.e., those at which ballast was discharged) were tallied. The historical data contained 15,628 unique voyages out of which 1,196 journeys had more than one port of ballast discharge. Out of 17,703 stops at ports of arrival, there were 6,287 incidents of reported ballast discharge. Note that the discharge may have presented an acceptable biosecurity risk, if the discharge were carried out in a regulated way, for example after open-ocean exchange. 'Intent to discharge' was not used as a predictor because there was poor agreement between the declared intention to discharge and the reported discharge volume. More information about this agreement can be found in the Suppl. material 1: Table S1 in the SM summarises the description of the variables used in the historical dataset for the ballast water analysis.

Biofouling

Biofouling on the submerged surfaces of 322 international merchant vessels was sampled by Inglis et al. (2010) upon arrival to New Zealand. This included 166 container and general cargo vessels, 49 passenger vessels, 37 roll-on roll-off (RoRo) / car carriers, 31 bulk carriers, 21refrigerated cargo ships, 12 tankers and 6 vessels of other miscellaneous class types. A standardised sampling design was used to collect samples of biofouling from the hull and external niche area surfaces of the vessels. A questionnaire administered to the vessel's master provided information on the design features of the vessel, hull maintenance record (e.g., age and type of antifouling coatings), and recent voyage history.

Data on the presence and mass of biofouling from 314 of the merchant vessels and associated data from the questionnaire were used to build the biofouling model. As the contemporary arrival data to which the model would predict contained only limited information on the recent voyage history of the vessels, the models were fitted using variables from the historic study that could be easily matched to contemporary information collected by BNZ. They included design specifications of the vessel (e.g., class type, design speed), vessel age, niche area, the age of antifouling coatings on arrival, and the maximum periods of inactivity (lay-up) in any location. For biofouling data, the main two response variables were biomass of biofouling per square metre and presence/absence of any biofouling. The latter had a value of 1 whenever biofouling was present, and zero if there were no biofouling. This binary variable was used to correct mass; that is, whenever mass had a zero value, but biofouling was present, a correction value of 0.001 g was added to the mass. Predictor variables in the historical biofouling data were the age of antifouling coating, vessel type, vessel age, vessel speed, season, niche, and maximum duration of lay-up. The description of the variables and their ranges are given in the Suppl. material 1: Table S2.

Data on contemporary shipping

Contemporary data sourced from BNZ on international vessel arrivals into New Zealand were a combination of two data sets. The first included data on vessel arrivals between January 2015 to December 2017, and was derived from mandatory pre-arrival documentation provided by the vessel master to BNZ (Ministry for Primary Industries 2019). These data were originally collected to aid in the tactical evaluation of biosecurity risks. The second data set included port-by-port arrivals by merchant vessels, including the ports of first arrival to New Zealand and coastwise voyages for the year 2016. These data were purchased from the Lloyd's Maritime Intelligence Unit, and used to verify the completeness of the BNZ data. Details of all data preparation for the analysis are provided in the Suppl. material 1 – part 1. The same contemporary data were used for both ballast water and biofouling analysis, but a different set of predictors were selected for each.

Model construction

The ballast discharge model was used to predict the total annual volume of ballast water discharged at each New Zealand port. The contemporary data lacked ballast discharge information, so we had to predict it from historical data. The models were built to predict ballast discharge at each port as a function of arrival port and ship characteristics.

The historical biofouling data from Inglis et al. (2010) were used to fit a model that predicted the presence and biomass of biofouling on vessels that arrived in New Zealand between 2015 and 2017. The total mass of biofouling on a vessel was estimated as the product of its estimated mass per unit niche area (i.e., density) and the total niche area

for the vessel. A model was constructed from historical data to predict the presence and per unit area of biofouling in niche areas of the vessels as a function of vessel characteristics such as the age of antifouling coatings on the vessel, and the frequency and duration of lay-up periods. As most biofouling occurs in niche areas (Inglis et al. 2010; Davidson et al. 2016), the total mass of biofouling on a vessel was estimated by calculating the total area of niches on each vessel using relationships developed by Moser et al. (2016) and scaling the predicted density for the entire vessel hull area. This model was then applied to the contemporary vessel data to predict port-level exposure to biofouling. Additional details about the modelling approach can be found in the Suppl. material 1 – part 3.

Ballast water

This section reports construction of an algorithm to predict the volume (and at which ports) a vessel will discharge their ballast water, given their vessel-level characteristics (e.g., type, DWT, Ballast capacity) and journey-level information. The complex aspect of this endeavour is that after entering New Zealand's waters, each vessel may visit more than one port, with the possibility of discharging ballast water at one or more of the ports visited. As a result of discharging at one or more ports, the variable of interest is structured as a vector for each vessel, so any algorithms investigated are required to take this vector-based outcome into account.

k-Nearest-Neighbours

An algorithm based on *k*-Nearest Neighbours (KNN, Fix and Hodges 1951) was developed to simultaneously predict which ports a vessel may visit and the subsequent discharge amount. For each journey in the contemporary data, this algorithm finds the *k* most similar journeys in the historical data based on some measure of similarity and uses the average ballast discharge from those journeys as a prediction of the discharge for the contemporary journey.

Denoted by X the n × p matrix of data used for model training (the training data is the historical data as described in the Suppl. material 1 – part 1), where n is the number of voyages (observations) and p is the number of fields (predictors) in the data; let Y be the corresponding n × v matrix of discharge volumes, where v is the number of ports. Let z_j be a vector of length p, containing the predictors for the j^{th} voyage in the contemporary data, for which we wish to predict ballast discharge,$\hat{y}_j$, at each of the ports. KNN algorithm as applied to ballast discharge data can be explained as below:

1. Repeat for $j = 1,\ldots,m$, where m is the number of voyages in the contemporary data for which we wish to predict:

   a) Calculate $D(z_j, X)$, the distance from z_j to each of the rows xi, $i = 1, \ldots, n$ in the training matrix X.

   b) Find S, the set of the k^{th} smallest distance in D (if there are ties, keep these in the set).

c) Calculate

$$\hat{y}_j = \frac{1}{|S|}\sum_{i \in S} y_i$$

i.e., the predicted discharge values are the average of the k^{th} nearest neighbours discharges.

2. For each year t in the contemporary data:
   a) Find $S_{t[v]}$, the set of all voyages arriving at port v during year t.
   b) Calculate

$$\hat{B}_{t[v]} = \sum_{i \in S_{t[v]}} \hat{y}_{t[j]}$$

i.e., $\hat{B}_{t[v]}$ is the total predicted ballast discharge in port v during year t.

The KNN algorithm does not work for categorical predictors, as there is no defined distance metric for mixtures of categorical and numerical predictors. To allow for categorical predictors (such as vessel type), the KNN algorithm was implemented on numerical predictor matrices created within each level of a category. For example, to use vessel type and dead weight tonnage as predictors, the algorithm was run on the matrix X formed from the subset of voyages from each level of vessel type until all predictions were made.

The historical data were structured such that each row of the dataset represented a single event in a vessel's voyage. For example, one row may consist of the arrival event, with no discharge, and the next row a discharge event for the same vessel at a different port. It was important that if a vessel did not discharge in a port, then this non-event was not recorded. In order to have a complete vector response for each voyage, it was reasonably assumed that non-events resulted in a 0 discharge being recorded. An example of this transformation is provided in Suppl. material 1: Fig. S2 in the Suppl. material 1 – part 4.

Choice of predictors and number of nearest neighbours, *k*

The choice of which predictors to include, and how many nearest neighbours to use in the prediction was made using cross-validation. Given our prediction target of total ballast water volume at a port within a year, our cross-validation strategy was to leave out a single year at a time to form the testing set, with the remaining year's data forming the training set. A schematic of cross-validation used for KNN analysis is given in Suppl. material 1: Fig. S3 in the Suppl. material 1 – part 4.

Model validation

In order to provide an unbiased validation of the chosen model and help understand its operational characteristics, the data were split into two: model-fitting data comprising all voyages within the years 1999 to 2005 (inclusive) and validation data comprising all voyages within the years 2006 and 2007.

Prediction errors

To estimate the prediction error from the KNN algorithm, a bootstrap procedure was used. Voyages were randomly sampled with replacement from the training data, and the KNN algorithm fit on each bootstrap sample. The standard deviation (and mean/median) of total discharge within a port was calculated from 100 bootstrap resampled values of total discharge.

Biofouling

Model selection

Several different modeling approaches were applied to the historical data in the process of model selection to estimate biofouling biomass for the contemporary dataset (Suppl. material 1 – part 6). Random forests provided the best model to obtain the prediction (Breiman 2001; Liaw and Wiener 2002; see Hastie et al. 2017 for a discussion) and we used quantile-regression random forests to obtain the distribution to represent the prediction uncertainty (Meinshausen 2017).

Random forests analysis was applied to the data because the algorithm is indifferent to the conditional distribution of the response variable, meaning that no special provisions need to be taken to allow for the heavy zero inflation. The random forest model explained about 15% of the variation in the biofouling mass, which was better than its alternatives. Quantile random forests was used to obtain journey-specific prediction distributions which could then be aggregated to obtain overall levels of uncertainty for the exposure of each port to biofouling for each year.

Model building for biofouling biomass

The likelihood of entry via biofouling was assumed to be proportional to the biofouling exposure at each port, each year. For a single vessel, this was assumed to be the predicted total mass of biofouling in the niche areas of the vessel (not including the hull as a niche area). However, the contemporary data did not contain biofouling mass measurements for each vessel but did contain measurements for each vessel's niche area in metres squared. To handle this, two models were constructed: a model to predict the biofouling mass per vessel as a function of several variables, and another model to predict the quantiles of the prediction distribution. The random forest model used the following continuous predictor variables: dry weight tonnage, days since last antifouling paint, niche area, and mean speed. Vessel type was assessed and was excluded from the models because the improvement of fit was insufficient.

To predict the density of biofouling for each vessel in the contemporary data, missing data were first handled using multivariate imputation via chained equations (Buuren and Groothuis-Oudshoorn 2010) to make five complete variant datasets, which

were analysed in parallel. The random forest biofouling presence model was applied to each contemporary vessel movement, which gives an estimate of the quantum of biofouling. Then the prediction distributions were estimated using the quantile random forest. Once the density of biomass present on each vessel was predicted, it was multiplied by the niche area of the vessel to give a total niche biomass prediction, which was then used in the allocation procedure.

Ranking ports by likelihood of marine NIS entry

We calculated an entry likelihood score from ballast discharge and biofouling propagule pressure on a scale of 0 (lowest) to 1 (highest). That is, the mean values from the estimated prediction distribution were scaled to each of the respective discharge and biofouling amounts by the maximum amount within the ports, which sets the scaled score to 1 for the port with the highest amount. We did not subtract the smallest so that no port was allocated a weight of 0. These scores were then averaged to give the relative entry likelihood score for each port. Weights were then calculated in proportion to these relative scores. The weights sum to 1 across the 15 ports and could, therefore, be used to allocate total surveillance effort among ports relative to the likelihood of entry of marine NIS for each port.

Results

Ballast discharge

We fitted the KNN model using all combinations of vessel type and arrival port as categorical variables, DWT or date of arrival as the nearest neighbour matrix, and k = 1,3,5 nearest neighbours to the model data set. Results of the cross-validation in Table 5 (Suppl. material 1 – part 7) showed that the KNN that grouped by both port of arrival and vessel type included DWT for calculating nearest neighbours, and using five nearest neighbours performed the best, i.e., this combination of distance variables had the smallest RMSE value (88732.8 for k = 5). A dataset of years 2006–2007 was used to test KNN validation by comparing the mean of predicted ballast water values (as predicted by the bootstrap) with the true values (Suppl. material 1: Fig. S18 in Suppl. material 1 – part 8). Based on the results of the KNN validation summarised in this Fig., the KNN models performed reasonably well for prediction; the observations follow the 1:1 line and double standard deviation lines for the port-level discharge mostly cross the line.

Predicting contemporary ballast discharge

The best model arising from the cross-validation fitting study of the KNN algorithm as selected was used to predict contemporary ballast discharge per port. For this application, the full historical data (1999–2007) were used for training. Fig. 1 shows the mean and 10% and 90% quantiles as estimated by bootstrapping. Tauranga, with an average of 2255.73 ±

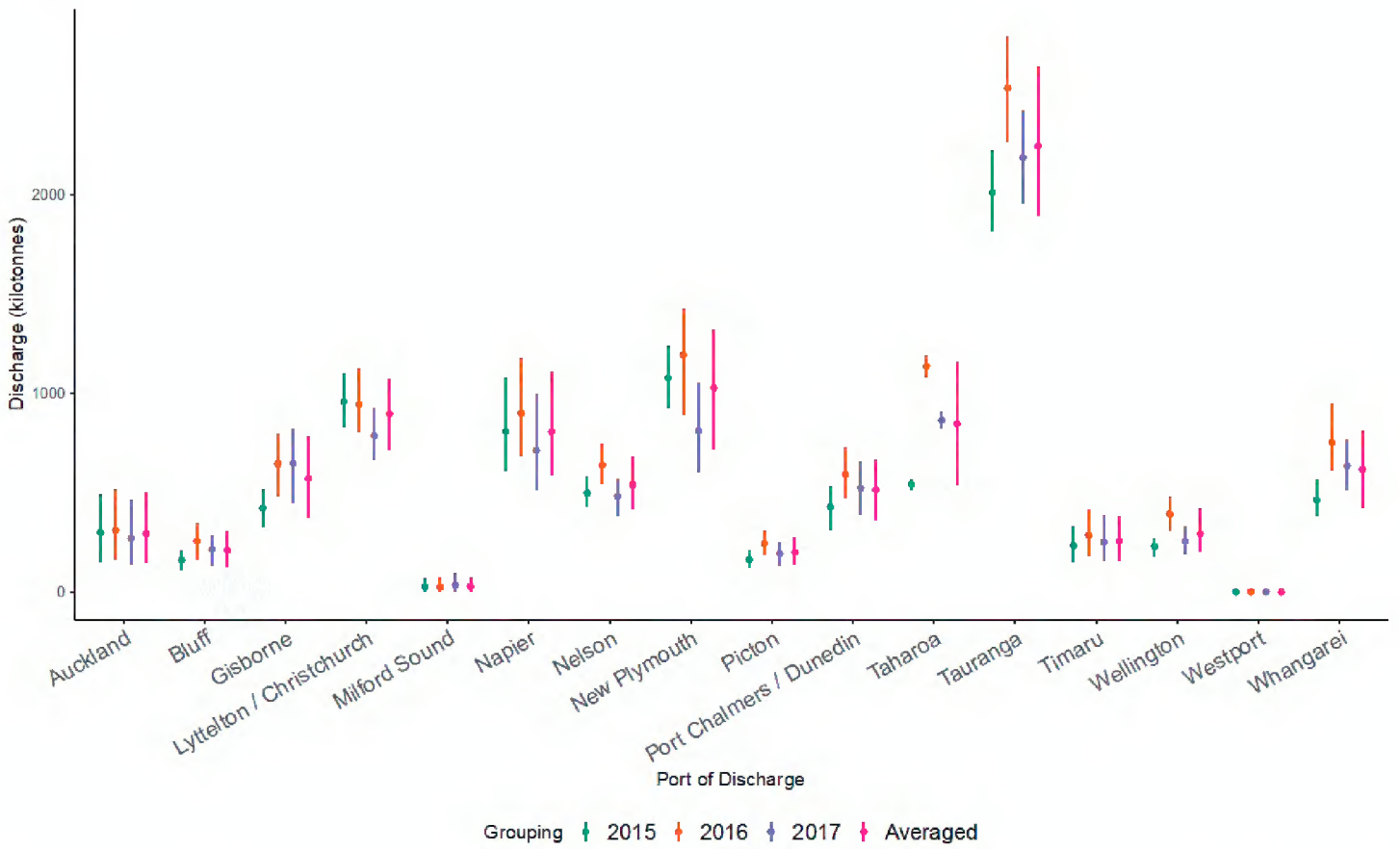

Figure 1. Predicted ballast water discharge (kilotonnes) per port for the contemporary data. Data are the mean and standard deviation of total volume as estimated by bootstrapping. The figure displays the mean and 10% and 90% quantiles of the bootstrap distribution.

292.86 kilotons, had the highest predicted average ballast discharge across the years 2015–2017, followed by New Plymouth (1025.22 ± 231.89), Lyttelton (907.44 ± 129.39), Taharoa (831.84 ± 242.45), Napier (812.54 ± 191.47), Whangarei (624.92 ± 165.55), and Gisborne (575.33 ± 160.74). For a port receiving a high number of vessels, Auckland had a low volume of discharge by comparison. Following an increase in 2016, the predicted ballast discharge decreased in 2017 in almost all ports (Fig. 1). The lowest predicted values for ballast discharge were observed in Milford Sound, Picton and Bluff, and zero for Westport.

Biofouling

Based on the Random Forest statistics, e.g., Gini impurity index and MSE (mean squared error), important variables for biofouling prediction were days since last antifouling paint, niche area, vessel speed and DWT. Variability in the modelling was accounted for by predicting quantiles of the vessel biomass (Fig. 2).

As summarised in Table 1, Auckland had the highest average predicted biofouling exposure, followed by Tauranga, Lyttelton, Napier, and Wellington. Ports with very low biofouling exposure were Whangarei, Gisborne, Picton, and Taharoa. Fig. 3 shows the variation of the actual biofouling and the prediction distribution, which demonstrates that the model is not of particularly good quality. That is, some of the data points, particularly for higher values of biofouling mass, are well below the 1:1 line, indicating an underestimation of higher values, which reflects regression to the mean.

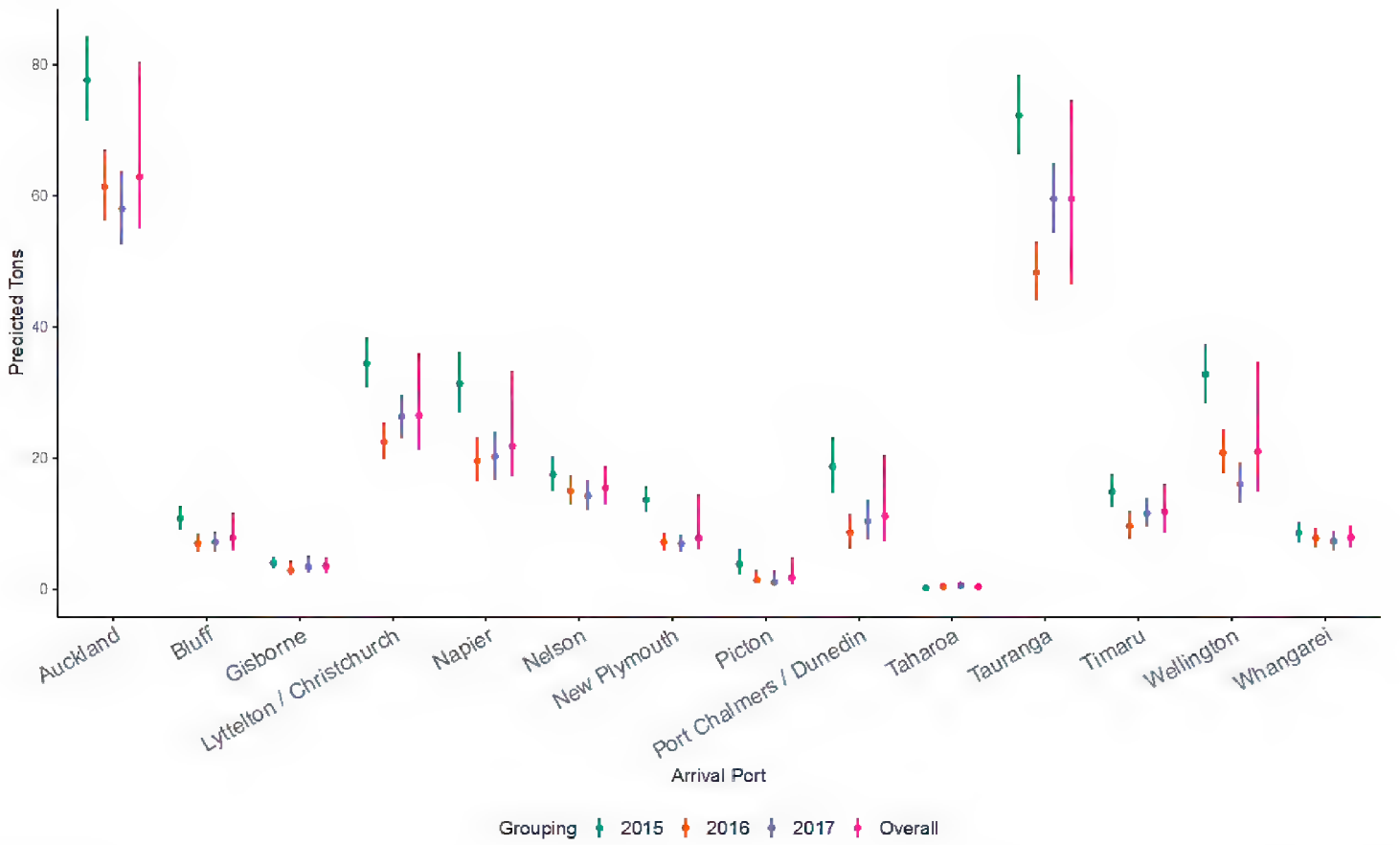

Figure 2. Predicted vessel biofouling exposure (tonnes) to each arrival port for the contemporary data (2015, 2016, 2017 and combined years). The median (dot) and 10% and 90% quantiles (upper and lower lines) of the prediction distribution are displayed.

Table 1. Predicted biofouling exposure and 10–90% quantile for the ports with the highest average biofouling mass as illustrated in Fig. 2.

Port	Predicted biofouling exposure (tonnes/year)	10–90% quantile (tonnes/year)
Auckland	70	58.4 – 86.02
Tauranga	65.37	49.62 – 79.82
Lyttelton	30.83	23.11 – 39.1
Napier	26.48	18.81 – 36.22
Wellington	25.81	15.44 – 37.48

Ballast water and biofouling exposure per port

Finally, for each arrival port, the biofouling exposure can be compared with the ballast discharge (Fig. 4). This shows that the arrival ports of Auckland and Tauranga have the highest exposure for biofouling, and Tauranga has the highest volume of ballast discharge (i.e., highest entry likelihoods). The next cluster of ports comprises New Plymouth, which had the second highest ballast discharge but received much lower biofouling exposure than many other ports; Wellington, which had low volume of ballast discharge and high exposure to biofouling; and Lyttelton and Napier, which seemed to have relatively high levels of exposure to both ballast discharge and biofouling. The balance of ports have relatively low exposure.

Based on the relative likelihood of entry of marine NIS via these two pathways, the highest surveillance weights would be allocated to the ports Tauranga and

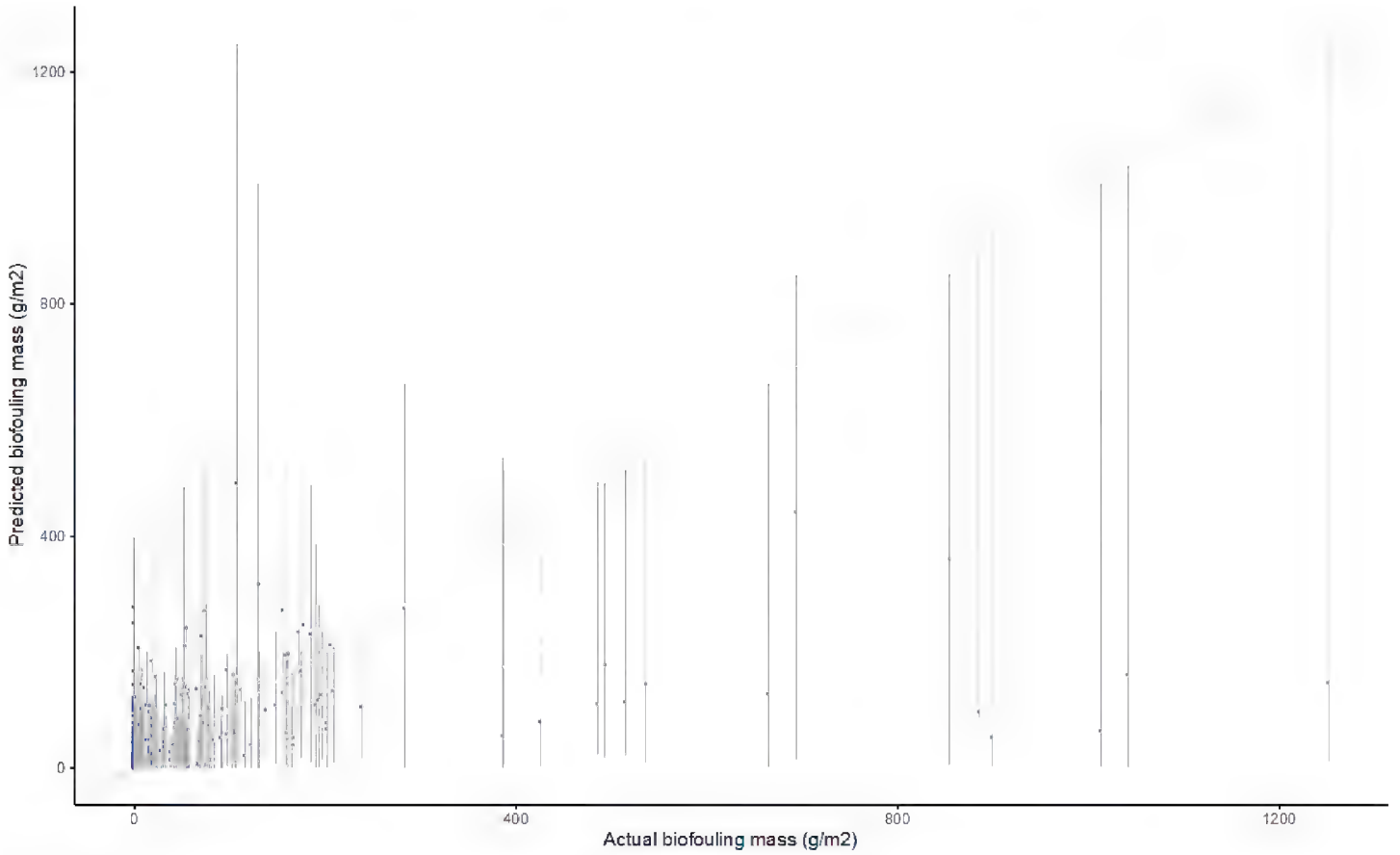

Figure 3. The variation of the actual biofouling mass (g/m2) versus the predicted biofouling mass (g/m2). The vertical grey lines represent the 10% and 90% quantiles of predicted values. The diagonal line shows the 1:1 relationship.

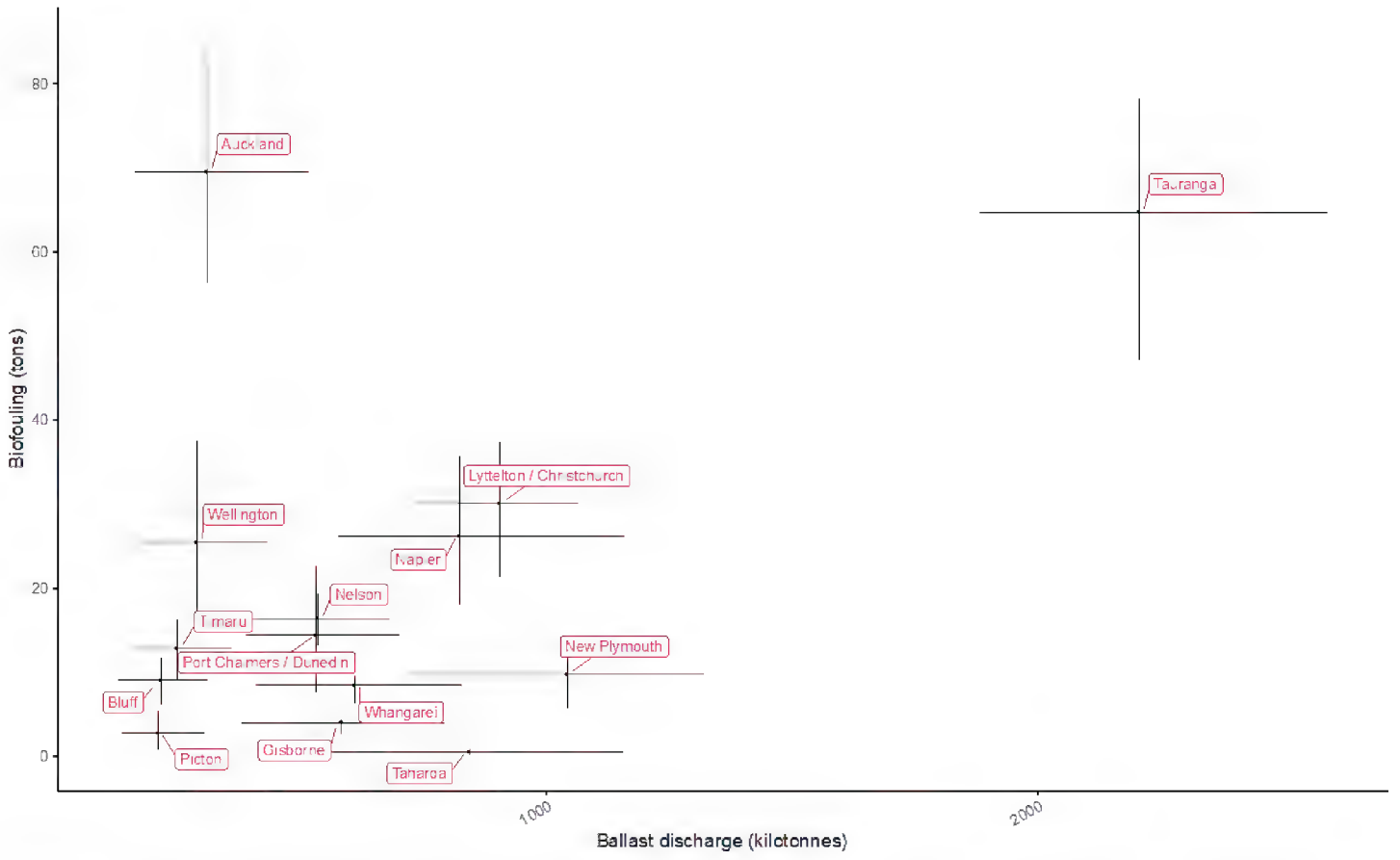

Figure 4. Predicted vessel ballast water discharge (kilotonnes) vs biofouling exposure (tonnes) per port. The figure displays the mean and 10% and 90% quantiles of the bootstrap distribution for ballast water and from the prediction distribution for biofouling. The values are the average of ballast water and biofouling mass over years 2015 – 2017.

Auckland, followed by Lyttelton, Napier, New Plymouth, and Wellington, whereas the lowest surveillance effort would be allocated to the ports Gisborne, Timaru, Bluff, and Picton.

Priority ports for marine surveillance

The relative entry likelihood weighting, calculated from the mean values of the estimated prediction distributions and scaled for the respective ballast discharge and biofouling exposure, is provided in Table 2. Based on these results, the ports with

Table 2. Relative entry likelihood for each port for ballast water discharge and biofouling exposure, and relative surveillance effort allocation weights.

Port	Total ballast discharge (kilotonnes)	Total biofouling exposure (tonnes)	Relative entry likelihood	Relative allocation weight
Tauranga	2,205,593	64.73	1	0.23
Auckland	307,807	69.53	0.59	0.13
Lyttelton	904,603	30.07	0.43	0.1
Napier	823,096	26.19	0.38	0.09
New Plymouth	1,042,600	9.80	0.31	0.07
Wellington	288,243	25.49	0.25	0.06
Nelson	534,271	16.38	0.24	0.06
Dunedin	530,440	14.45	0.23	0.05
Whangarei	609,168	8.51	0.2	0.05
Taharoa	842,739	0.54	0.2	0.05
Gisborne	581,713	4.03	0.16	0.04
Timaru	247,965	12.82	0.15	0.04
Bluff	213,933	9.05	0.11	0.03
Picton	208,567	2.76	0.06	0.02

higher likelihood of marine NIS entry were Tauranga, Auckland, Lyttelton, Napier, and New Plymouth with scores of 1, 0.59, 0.43, 0.38, and 0.31, respectively. Tauranga was assigned a relative allocation weight of 0.23, followed by Auckland (0.13), Lyttelton (0.1), and Napier (0.09). New Plymouth, Wellington, Nelson, Dunedin, Whangarei, and Taharoa were ranked next based on their surveillance allocation weights. Gisborne, Timaru, Bluff, and Picton had the lowest relative weighted scores. There was a positive relationship between the number of vessel arrivals per port and the entry likelihood scores (r^2 = 0.75). Ports where the entry likelihood is greater than might be predicted by numbers of arrivals alone include Tauranga, Napier, New Plymouth and Taharoa, where there are relatively large proportions of bulk carriers or tankers. Ports where entry likelihood seems lower include Auckland and Wellington. To investigate the variability of the ranking of sites, this allocation was also performed to each of the bootstrap replicates. Fig. 5 shows the histogram of ranks from this procedure.

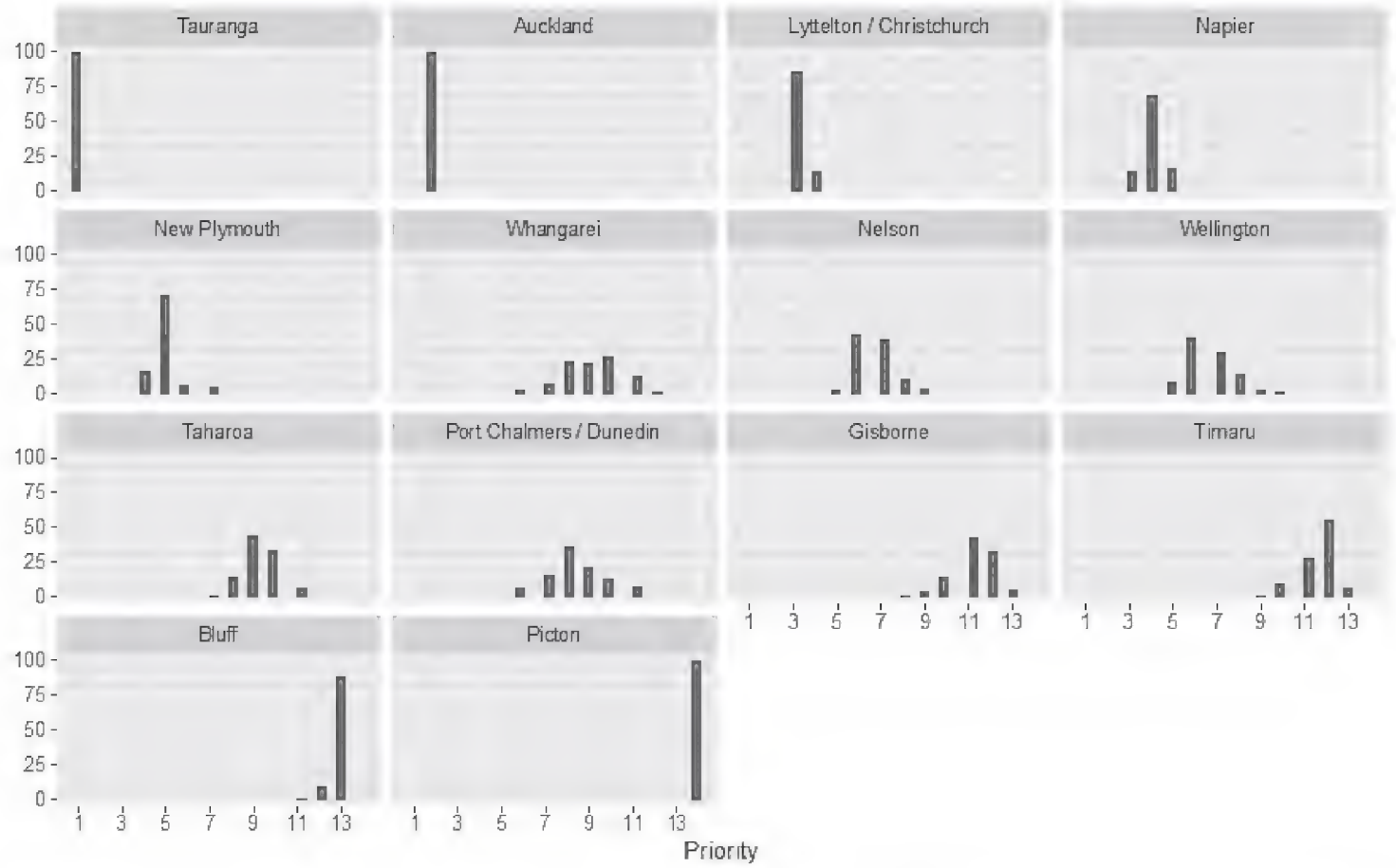

Figure 5. Rank histograms for surveillance allocation, as estimated by bootstrap and prediction simulation.

Discussion

We developed a systematic methodology that can be used to adapt surveillance effort over time in response to changes in propagule pressure. The methods can be applied to port systems in other countries or to other invasion pathways. Summarised at the highest level, the study applied model-based estimation to activity data, where the activity data were contemporary transits of marine vessels. In the case of ballast discharge, the model-based estimation was non-parametric, using KNN to identify the vessel transit trajectories that were most similar to the contemporary records on the available subset of fields. In a sense, the contemporary transit data reweighted the historical ballast discharge records, and the significant changes in activity observed in Tauranga were reflected in its enhanced likelihood of marine NIS entry via this pathway. For the biofouling exposure, we predicted vessel-level biofouling using random forests, and its uncertainty using quantile random forests, and computed port exposure directly from the contemporary transit records. Key in each case was the availability of contemporary activity data that could be used to calibrate the relative likelihood of marine NIS entry experienced by each port. In this system as in many others, the likelihood of entry scales smoothly with the exposure.

Based on this study, Tauranga had the highest predicted joint ballast discharge volume and exposure to biofouling. This is despite Auckland receiving the highest number of vessel visits in almost all years (Suppl. material 1: Fig. S5). It has previously been shown that total number of visits alone is not a reliable predictor of propagule pressure (Verling et al. 2005), and our results reinforce this outcome. Although the busiest

ports – such as Auckland, Tauranga, and Lyttelton – had the highest relative exposure to ballast discharge and biofouling, other highly visited ports such as Wellington and Nelson had lower levels when compared to less visited ports such as Napier and New Plymouth. Several studies have relied on vessel visits as a proxy for propagule pressure in ballast water (Drake and Lodge 2004) and this might be a source of error due to significant differences between vessel types and their characterises and operational profiles (Minton et al. 2005; Verling et al. 2005). For example, Carney et al. (2017) reported that the increase in annual ballast discharge was not because of an increase in total vessel arrival numbers, but instead resulted from an increase in bulk carrier traffic. Thus, it seems that considering vessel type, size and discharge capacity will provide a better predicator of ballast discharge volume for the vessel.

Higher entry likelihood at Tauranga is likely related to this port having the highest number of discharge declarations, vessel visits and DWT. Tauranga was also visited with the highest number of bulk vessels, which along with containers had the highest discharge counts in the historical data. Considering that bulk vessels had the highest DWT and discharge volume among the vessel types, it is not surprising that their size and discharge volume contributed to their higher likelihood of marine NIS entry in Tauranga. Increase in ballast discharge in Tauranga, which reached its peak in 2016, was accompanied by the highest number of bulk carriers and containers visiting this port in 2016. Predicted ballast discharges were higher in 2016 mainly because the total annual number of vessels visiting New Zealand ports was highest in 2016; all the vessel types except for general cargo visited New Zealand's ports more in 2016 and had a higher DWT in this year. The total number of visits decreased in 2017, along with the numbers of visits by almost all vessel types including those with high DWT. This can explain the decrease in ballast water predicted values in 2017.

In several studies, ballast discharge volume and frequency have been reported to vary significantly by ship type (Verling et al. 2005; Minton et al. 2015), with the largest volumes of ballast water from bulkers and tankers (McGee et al. 2006; Cordell et al. 2009; Carney et al. 2017). Vessel type was consistently suggested as a reliable predictor of biofouling, as different vessels vary in the size and number of niche areas (Lacoursière-Roussel et al. 2012; Lane et al. 2018). Based on the results of cross-validation and model fit statistics, vessel type was not included as a variable in our models, but the variation may have been more efficiently represented by the included predictors, e.g., DWT. In this study, niche area and days since antifouling paint were the two most important variables explaining the variation of biofouling mass. Time since last cleaning and last antifouling paint also played a role in the increased likelihood of entry of marine NIS from biofouling (Floerl and Inglis 2005; Drake and Lodge 2007; Lacoursière-Roussel et al. 2012). Biofouling mass on the vessel might also vary depending on the other factors such as vessel speed as lower speeds and inactivity appear to give the fouling communities opportunity to settle (Minchin and Gollasch 2003).

As noted earlier, unregulated ballast discharge is an important pathway for marine NIS spread. Consequently, information about ballast discharge quantities, along with detailed information on ballast discharge mitigation measures such as open-ocean ex-

change, would be valuable for risk analysis. This information would enable reporting of potentially risky discharge events, which could then be used to inform the ports that were most in need of surveillance. However, such records were not available for the contemporary data, and only intention-to-discharge records were available in the historical data.

This study predicted the amount of discharged ballast water and biofouling mass per port at the voyage level, and the ports where the discharge occurred. The predicted values of discharged ballast water and biofouling mass were used to review New Zealand's marine biosecurity surveillance programme by estimating the relative likelihood of entry of marine NIS at each port and allocating surveillance effort between ports accordingly. Some studies have aimed to identify hot spots of marine NIS exchange via ballast water (Drake and Lodge, 2004), or possible source ports and countries for the arrival of a specific NIS by analysing international shipping networks and generating pathway simulations to help optimise shipping container inspection protocols (Paini and Yemshanov 2012). This study differs from previous studies in that a combination of information about vessel traffic patterns and vessel characteristics was used to identify ports likely to receive marine NIS, whilst including a measure of propagule pressure associated with each vessel arrival. Other studies tried to prioritise locations for tracking biological invasions, but used less exhaustive analyses and were more descriptive to illustrate the utility of the database (e.g., Molnar et al. 2008) or to prioritise the pathways that pose the greatest threat (e.g., McGee et al. 2006). In comparison, we developed a systematic methodology that can be used to adapt surveillance effort over time in response to changing vessel profiles and can be applied to ports in other countries or globally. The techniques used in this study allowed us to consider multiple discharges during a voyage for ballast water analysis, and to account for biofouling presence and mass while predicting voyage-level, port-level, and annual propagule pressure.

This study has limitations. We did not incorporate other factors that might mitigate or exacerbate entry likelihood, such as the efficiency of ballast water exchange or treatment in removing marine NIS (Minton et al. 2005), or the influence of transit time on ballast water organism survival prior to discharge (Cordell et al. 2009). Our analysis gave equal weighting to the importance of ballast water and biofouling as pathways for the introduction of marine NIS. Although numerous studies have attempted to summarise the relative importance of these and other pathways for the introduction of marine NIS based on records of established species, such studies typically include many historical (e.g., pre-1950) introduction records, hence the numbers of marine NIS attributed to each pathway may not reflect contemporary patterns of transport. Moreover, many marine NIS can be transported within ballast water and as biofouling, making it difficult to distinguish the relative contribution of each pathway (Fofonoff et al. 2003).

Drake and Lodge (2007) estimated the abundance of fouling organisms on a typical merchant vessel to be between 5.85×10^5 and 6.19×10^6 individuals, which is within the range of abundance of organisms contained in untreated ballast discharges ($2.63 \times 10^5 - 9.83 \times 10^6$ individuals; Minton et al. (2005)). While these gross metrics of per vessel propagule pressure are comparable, it is unclear whether

organisms are equally likely to be introduced, or if introductions by each pathway have the same likelihood of survival and establishment. In the absence of this information the standardised pathway weightings seem appropriate. In aligning surveillance with contemporary patterns of international shipping it is also important to consider other influences on the allocation of survey effort. For example, our analysis did not include port arrivals by small ("recreational") craft, which also carry burdens of biofouling (Floerl et al. 2008; Lane et al. 2018). More than 600 small marine craft enter New Zealand waters each year, with around two thirds of these clearing Customs in Opua. Other significant places of first arrival for small craft are Auckland, Whangarei and Picton which, with Opua, account for more than 90% of annual arrivals. We also did not consider the relationship between surveillance effort and detection probability at the scale of the port, which can be influenced by the size and dispersion of habitats suitable for establishment by marine NIS. For example, the total area of Waitematā Harbour, Auckland, is at least 3 × that of Nelson Harbour and contains the Port of Auckland and 11 marinas for small craft which are distributed from the mouth to the head of the harbour (Morrisey et al. 2012). Gaining equitable survey sensitivity across the surveyed ports requires a proportionately larger sample effort in larger harbours.

Specific policy outcomes regarding surveillance are beyond the scope of this paper. Nonetheless, the change in the pattern of vessel transits between the early 2000's and the contemporary data show that the propagule pressure has sharply increased in Tauranga relative to the other ports, and is now similar to that of Auckland. Further, Napier is subject to similar levels of propagule pressure to locations currently included in the MHRSS programme.

Conclusions

We developed a systematic methodology that can be used to adapt surveillance effort over time in response to changing vessel arrival patterns and types, and that can be applied to maritime ports in other countries or to other invasion pathways. Aligning survey effort with marine NIS entry likelihoods will increase the likelihood of early detection and improve management outcomes. The systematic likelihood-based methodology designed here is flexible and can be applied to surveillance programmes at any time if changes to propagule pressure occur.

Acknowledgements

The authors acknowledge the financial and other support provided by the New Zealand Ministry for Primary Industries, the Australian Department of Agriculture, Water and the Environment, and the University of Melbourne through the Centre of Excellence for Biosecurity Risk Analysis, an initiative of the Australian government.

References

Bell A, Phillips S, Georgiades E, Kluza D (2011) Risk analysis: vessel biofouling. MAF Wellington, 151 pp.

Breiman L (2001) Random forests. Machine Learning 45(1): 5–32. https://doi.org/10.1023/A:1010933404324

Buuren SV, Groothuis-Oudshoorn K (2010) MICE: Multivariate imputation by chained equations in R. Journal of Statistical Software 45: 1–68. https://doi.org/10.18637/jss.v045.i03

Carney KJ, Minton MS, Holzer KK, Miller AW, McCann LD, Ruiz GM (2017) Evaluating the combined effects of ballast water management and trade dynamics on transfers of marine organisms by ships. PLoS ONE 12(3): e0172468. https://doi.org/10.1371/journal.pone.0172468

Clarke S, Hollings T, Liu N, Hood G, Robinson AP (2017) Biosecurity risk factors presented by international vessels: A statistical analysis. Biological Invasions 19(10): 2837–2850. https://doi.org/10.1007/s10530-017-1486-1

Cope RC, Prowse TA, Ross JV, Wittmann TA, Cassey P (2015) Temporal modelling of ballast water discharge and ship-mediated invasion risk to Australia. Royal Society Open Science 2(4): e150039. https://doi.org/10.1098/rsos.150039

Cordell JR, Lawrence DJ, Ferm NC, Tear LM, Smith SS, Herwig RP (2009) Factors influencing densities of non-indigenous species in the ballast water of ships arriving at ports in Puget Sound, Washington, United States. Aquatic Conservation 19(3): 322–343. https://doi.org/10.1002/aqc.986

Cranfield HJ, Gordon DP, Willan RC, Marshall BA, Battershill CN, Francis MP, Nelson WA, Glasby CJ, Read CB (1998) Adventive marine species in New Zealand, 48 pp.

Davidson I, Scianni C, Hewitt C, Everett R, Holm E, Tamburri M, Ruiz G (2016) Mini-review: Assessing the drivers of ship biofouling management – aligning industry and biosecurity goals. Biofouling 32(4): 411–428. https://doi.org/10.1080/08927014.2016.1149572

Drake JM, Lodge DM (2004) Global hot spots of biological invasions: Evaluating options for ballast-water management. Proceedings of the Royal Society of London. Series B, Biological Sciences 271(1539): 575–580. https://doi.org/10.1098/rspb.2003.2629

Drake JM, Lodge DM (2007) Hull fouling is a risk factor for intercontinental species exchange in aquatic ecosystems. Aquatic Invasions 2(2): 121–131. https://doi.org/10.3391/ai.2007.2.2.7

Fix E, Hodges JL (1951) Discriminatory analysis - nonparametric discrimination: consistency properties. Report #4 US Air Force School of Aviation Medicine, Random Field, Texas. Published in Silverman and Jones (1989). https://doi.org/10.1037/e471672008-001

Floerl O, Inglis GJ (2005) Starting the invasion pathway: The interaction between source populations and human transport vectors. Biological Invasions 7(4): 589–606. https://doi.org/10.1007/s10530-004-0952-8

Floerl O, Smith M, Inglis G, Davey N, Seaward K, Johnston O, Fitridge I, Rush N, Middleton C, Coutts A (2008) Vessel biofouling as a vector for the introduction of non-indigenous marine species to New Zealand: Recreational yachts. NIWA report CHC2006-024.

Fofonoff PW, Ruiz GM, Steves B, Carlton JT (2003) In ships or on ships? Mechanisms of transfer and invasion for nonnative species to the coasts of North America. In: Ruiz GM,

Carlton JT (Eds) Invasive species: vectors and management strategies. Island Press, Washington, 152–182.

Georgiades E, Kluza D, Bates T, Lubarsky K, Brunton J, Growcott A, Smith T, McDonald S, Gould B, Parker N, Bell A (2020) Regulating vessel biofouling to support New Zealand's marine biosecurity system-a blue print for evidence-based decision making. Frontiers in Marine Science 7: e7. https://doi.org/10.3389/fmars.2020.00390

Hastie T, Robert T, Jerome F (2017) The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, 764 pp.

Hewitt C, Campbell ML (2010) The relative contribution of vectors to the introduction and translocation of invasive marine species.

Hewitt, CL, Campbell ML (2010) The relative contribution of vectors to the introduction and translocation of invasive marine species. Prepared for the Department of Agriculture, Fisheries and Forestry (DAFF), Australia.

Hewitt CL, Gollasch S, Minchin D (2009) The vessel as a vector-biofouling, ballast water and sediments. Biological invasions in marine ecosystems. Springer, 117–131. https://doi.org/10.1007/978-3-540-79236-9_6

Hulme PE (2009) Trade, transport and trouble: Managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46(1): 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Inglis GJ (2001a) Criteria for identifying and selecting high value locations and locations at risk of invasion by exotic marine organisms in New Zealand. Final research report for Ministry of Fisheries research project ZBS2000/01A, objectives 1 & 2. National Institute of Water and Atmospheric Research, Wellington, 27 pp.

Inglis GJ (2001b) Criteria for selecting New Zealand ports and other points of entry that have a high risk of invasion by new exotic marine organisms. National Institute of Water and Atmospheric Research, Wellington, New Zealand, 31 pp.

Inglis GJ, Hurren H, Gust N, Oldman J, Fitridge I, Floerl O, Hayden B (2006) Surveillance design for early detection of unwanted exotic marine organisms in New Zealand., Prepared for MAF Biosecurity New Zealand Post-clearance Directorate for Project ZBS2001-01, Wellington.

Inglis GJ, Floerl O, Ahyong S, Cox S, Unwin M, Ponder-Sutton A, Seaward K, Kospartov M, Read G, Gordon D, Hosie A, Nelson W, d'Archino R, Bell A, Kluza D (2010) The biosecurity risks associated with biofouling on international vessels arriving in New Zealand: summary of the patterns and predictors of fouling. Prepared for Biosecurity New Zealand Policy and Risk Directorate, National Institute of Water and Atmospheric Research. NIWA, Wellington.

International Maritime Organization [IMO] (2017) New Global Project to Address Bioinvasions Via Ship's Hulls. International Maritime Organization London.

Kolar CS, Lodge DM (2001) Progress in invasion biology: Predicting invaders. Trends in Ecology & Evolution 16(4): 199–204. https://doi.org/10.1016/S0169-5347(01)02101-2

Kospartov K, Inglis GJ, Seaward K, van den Brink A, D'Archino R, Ahyong S (2008) Non-indigenous and cryptogenic marine species in New Zealand. MAF Biosecurity New Zealand Post-clearance Directorate for Project BNZ10740.

Lacoursière-Roussel A, Forrest BM, Guichard F, Piola RF, McKindsey CW (2012) Modeling biofouling from boat and source characteristics: A comparative study between Canada and

New Zealand. Biological Invasions 14(11): 2301–2314. https://doi.org/10.1007/s10530-012-0230-0

Lane SE, Hollings T, Hayes KR, McEnnulty FR, Green M, Georgiades E, Robinson AP (2018) Risk factors for fouling biomass: Evidence from small vessels in Australia. Biofouling 34(9): 1032–1045. https://doi.org/10.1080/08927014.2018.1536202

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2: 18–22.

McCarthy SA, Khambaty FM (1994) International dissemination of epidemic Vibrio cholerae by cargo ship ballast and other nonpotable waters. Applied and Environmental Microbiology 60(7): 2597–2601. https://doi.org/10.1128/aem.60.7.2597-2601.1994

McGee S, Piorkowski R, Ruiz G (2006) Analysis of recent vessel arrivals and ballast water discharge in Alaska: Toward assessing ship-mediated invasion risk. Marine Pollution Bulletin 52(12): 1634–1645. https://doi.org/10.1016/j.marpolbul.2006.06.005

Meinshausen N (2017) QuantregForest: Quantile Regression Forests. R package version 1.3–7. R Foundation for Statistical Computing.

Minchin D, Gollasch S (2003) Fouling and ships' hulls: How changing circumstances and spawning events may result in the spread of exotic species. Biofouling 19(sup1): 111–122. https://doi.org/10.1080/0892701021000057891

Ministry for Primary Industries (2019) Arrival Process Steps - Vessels. Ministry for Primary Industries, Wellington, New Zealand.

Minton MS, Verling E, Miller AW, Ruiz GM (2005) Reducing propagule supply and coastal invasions via ships: Effects of emerging strategies. Frontiers in Ecology and the Environment 3(6): 304–308. https://doi.org/10.1890/1540-9295(2005)003[0304:RPSACI]2.0.CO;2

Minton MS, Miller AW, Ruiz GM (2015) Implications of Ship Type on Delivery and Management of Ballast Water. Biological invasions in changing ecosystems. Sciendo Migration: 343–364. https://doi.org/10.1515/9783110438666-021

Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment 6(9): 485–492. https://doi.org/10.1890/070064

Morrisey D, Inglis GJ, Peacock L, Seaward K (2012) Stochastic scenario tree modelling for the Marine High Risk Surveillance. Prepared for the Ministry for Primary Industries (SOW12099). NIWA. 1 Client report, Nelson.

Moser CS, Wier TP, Grant JF, First MR, Tamburri MN, Ruiz GM, Miller AW, Drake LA (2016) Quantifying the total wetted surface area of the world fleet: A first step in determining the potential extent of ships' biofouling. Biological Invasions 18(1): 265–277. https://doi.org/10.1007/s10530-015-1007-z

Paini DR, Yemshanov D (2012) Modelling the arrival of invasive organisms via the international marine shipping network: A Khapra beetle study. PLoS ONE 7(9): e44589. https://doi.org/10.1371/journal.pone.0044589

Seaward K, Acosta H, Inglis GJ, Wood B, Riding TA, Wilkens S, Gould B (2015) The Marine Biosecurity Porthole-a web-based information system on non-indigenous marine species in New Zealand. Management of Biological Invasions : International Journal of Applied Research on Biological Invasions 6(2): 177–184. https://doi.org/10.3391/mbi.2015.6.2.08

Silverman BW, Jones MC (1989) E. Fix and J. L. Hodges (1951): An inportant contribution to nonparametric discrimination analysis and density estimation. International Statistical Review 57: 233–247. https://doi.org/10.2307/1403796

Tauxe RV, Mintz ED, Quick RE (1995) Epidemic cholera in the new world: Translating field epidemiology into new prevention strategies. Emerging Infectious Diseases 1(4): 141–146. https://doi.org/10.3201/eid0104.950408

Verling E, Ruiz GM, Smith LD, Galil B, Miller AW, Murphy KR (2005) Supply-side invasion ecology: Characterizing propagule pressure in coastal ecosystems. Proceedings. Biological Sciences 272(1569): 1249–1257. https://doi.org/10.1098/rspb.2005.3090

Supplementary material 1

Supplementary materials

Authors: Rezvan Hatami, Graeme Inglis, Stephen E. Lane, Abraham Growcott, Daniel Kluza, Catherine Lubarsky, Charlotte Jones-Todd, Kimberley Seaward, Andrew P. Robinson

Data type: Docx file.

Explanation note: This is a file that contains the supplementary analyses.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.72.77266.suppl1